aws Webinar

Deep Dive on Amazon EFS

09/25/2018

Darryl S. Osborne

Solutions Architect – Amazon File Services

Your journey to Amazon Elastic File System, in four phases

Phase 1:
Choose the right storage solution

Phase 2:
Test and optimize

Phase 3:
Ingest

Phase 4:
Run

Phase 1:

Choose the right storage solution

What do you think about when choosing a storage solution?

Storage type

Features
and performance

Economics

Three types of storage

File

Block

Object

Three types of storage

File

Data stored as files in a directory hierarchy

Shared over a network

Three types of storage

Block

Data stored as blocks on a disk or disks

Locally attached

Three types of storage

Object

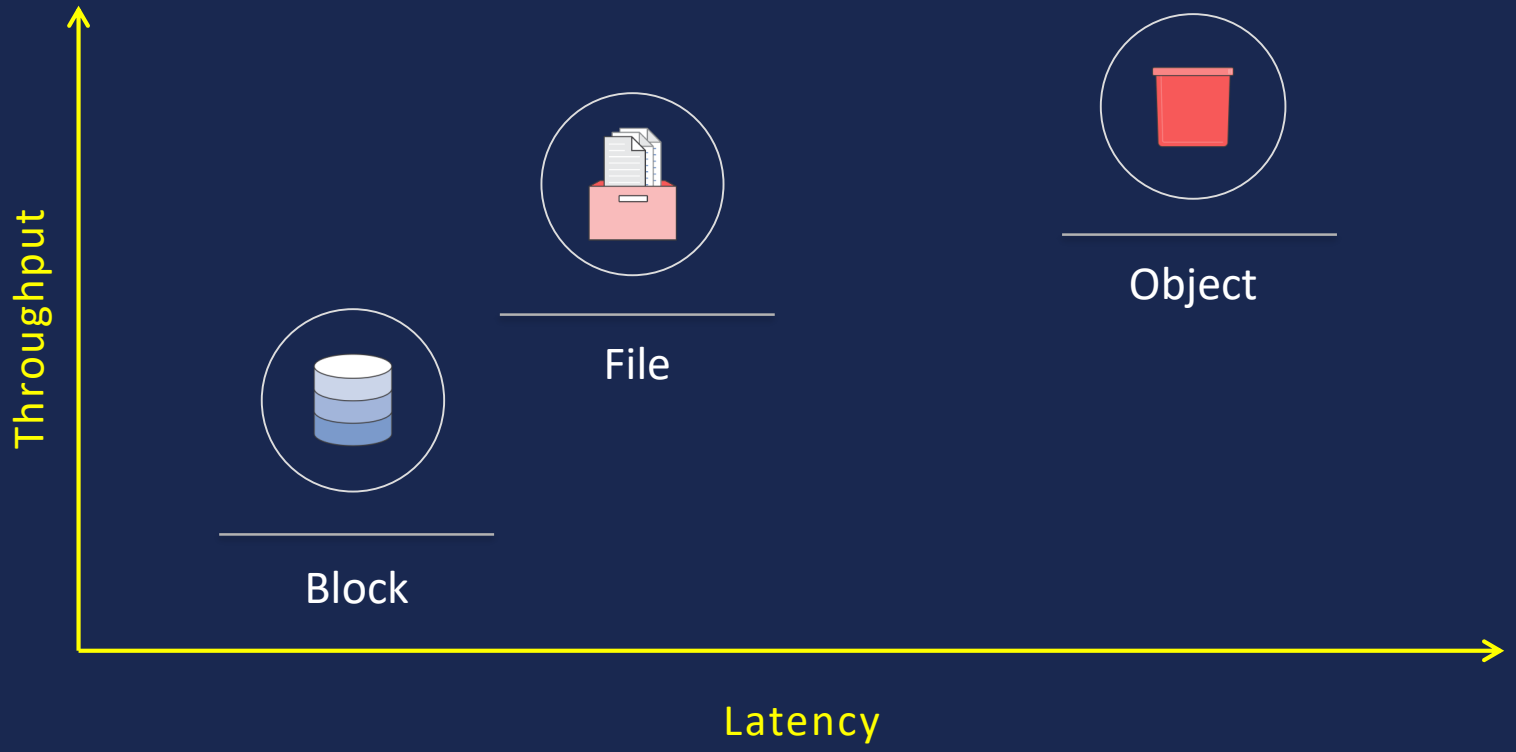
Data is stored as an object that's identified by a key in a flat space

Simple API to get and put data based on key

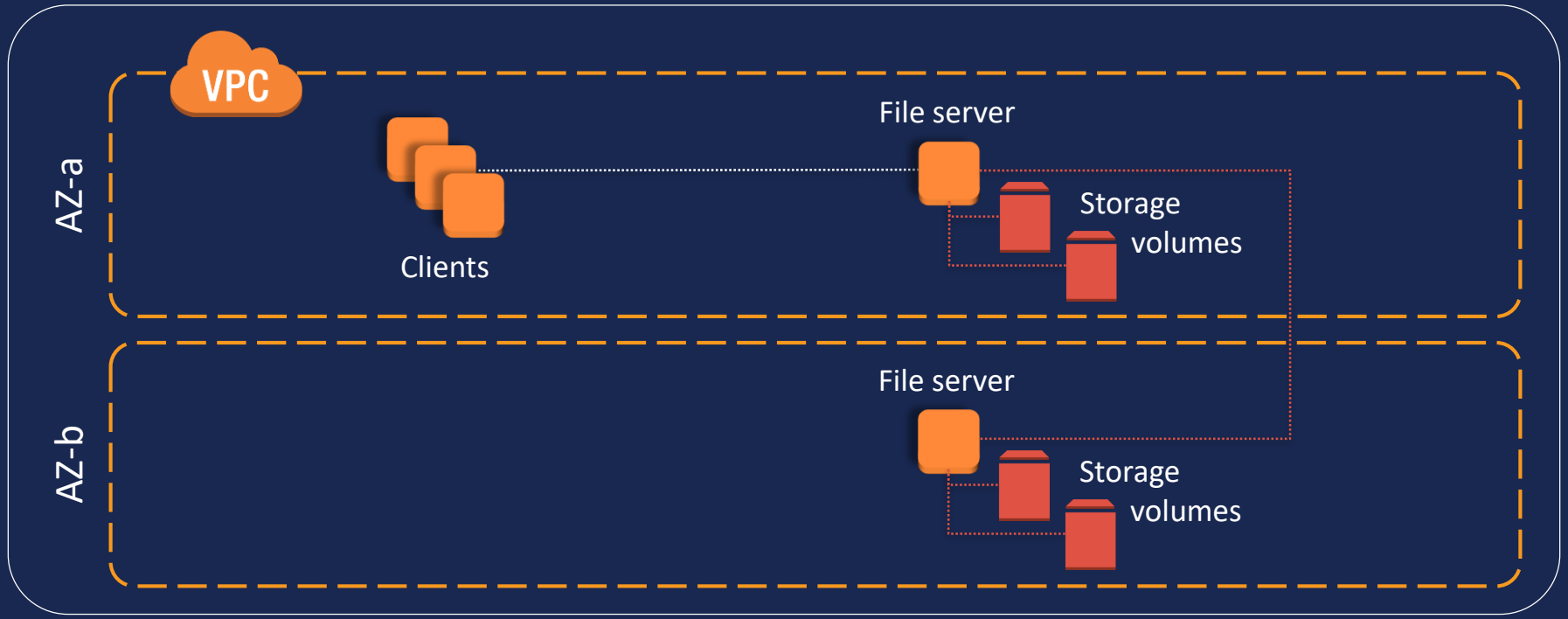
Why is file storage so popular?

- Works natively with operating systems
- Provides shared access while providing consistency guarantees and locking functionality
- Provides hierarchical namespace

How does performance compare



Before Amazon EFS… DIY file storage costs



Amazon EFS

A fully managed file service

aws Webinar

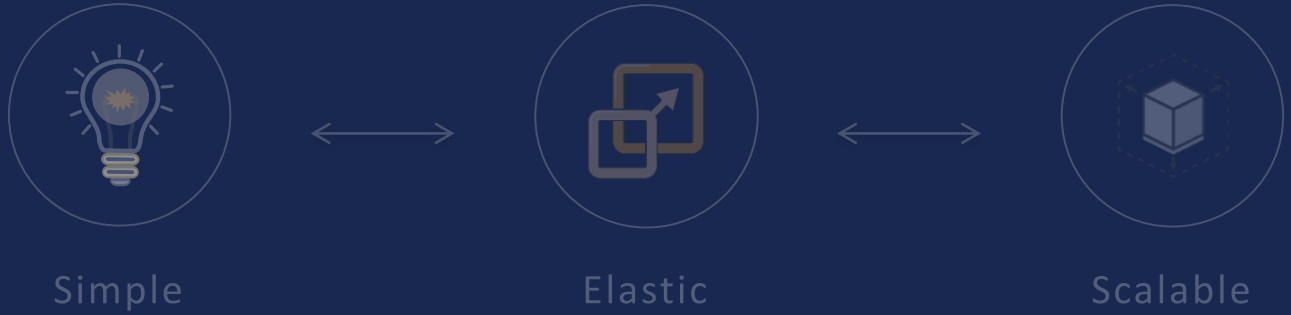
Key features of Amazon EFS

Simple

Key features of Amazon EFS

Key features of Amazon EFS

Key features of Amazon EFS



Highly available and durable

Key features of Amazon EFS

Simple

Elastic

Scalable

Highly available and durable

Performance modes

Performance modes

General Purpose (default)

Recommended for the majority of workloads

Performance modes

	General Purpose (default)	Max I/O
What it is for	Latency-sensitive applications and general-purpose workloads	Large-scale and data-heavy applications
Advantages	Lowest latencies for file operations	Virtually unlimited ability to scale out throughput/IOPS
Trade-offs	Limit of 7k ops/sec	Slightly higher latencies
When to use	Best choice for most workloads	Consider for large scale-out workloads

Recommended for the majority of workloads

Recommended for scale out workloads

Throughput modes

Throughput modes

Bursting Throughput
(default)

Recommended for the
majority of workloads

Throughput modes

Bursting Throughput
(default)
Recommended for the majority of workloads

Provisioned Throughput
Recommended for higher throughput to storage ratio workloads

Throughput modes

	Bursting Throughput (default)	Provisioned Throughput
What it is for	Varying throughput workloads	Higher-consistent throughput workloads
		
Advantages	Auto-scaling throughput	User-defined throughput
Trade-offs	Fixed throughput to storage ratio	Separate throughput charge
	Recommended for the majority of workloads	Recommended for higher throughput to storage ratio workloads
When to use	Best choice for most workloads	Ingest or higher throughput to storage ratio

Provisioned Throughput mode

Provisioned Throughput mode

Independent throughput

Provision throughput independent of data stored

Provisioned Throughput mode

Independent throughput

Provision throughput independent of data stored

Increase

As often as you need

Provisioned Throughput mode

Independent throughput

Provision throughput independent of data stored

Increase

As often as you need

Switch or decrease

Once every 24+ hours

Security model & compliance

Security model & compliance

Control network traffic using Amazon VPC security groups and network ACLs

Security model & compliance

Control network traffic using Amazon VPC security groups and network ACLs

Control file and directory access using POSIX permissions

Security model & compliance

Control network traffic using Amazon VPC security groups and network ACLs

Control file and directory access using POSIX permissions

Control administrative access (API access) using AWS IAM (action-level and resource-level permissions)

Security model & compliance

Control network traffic using Amazon VPC security groups and network ACLs

Control file and directory access using POSIX permissions

Control administrative access (API access) using AWS IAM (action-level and resource-level permissions)

Encrypt data at rest in transit

Security model & compliance

Control network traffic
using Amazon VPC security groups and network ACLs

Control file and directory access
using POSIX permissions

Control administrative access (API access)
using AWS IAM (action-level and resource-level permissions)

Encrypt data
at rest
in transit

Achieve Compliance
HIPAA-eligible
BAA
PCI DSS

Where is Amazon EFS available today?

- **US West** (Oregon)
- **US West** (N. California)
- **US East** (N. Virginia)
- **US East** (Ohio)
- **EU** (Ireland)
- **EU** (Frankfurt)
- **Asia Pacific** (Sydney)
- **Asia Pacific** (Seoul)
- **Asia Pacific** (Tokyo)
- **Asia Pacific** (Singapore)

More coming soon!

Amazon EFS economics

No minimum commitments
or upfront fees

No need to provision
storage in advance

No other fees
or charges

Bursting Throughput mode pricing

Single pricing dimension

Storage price

Pay only for the amount of storage you use per month

Includes 50 KiB/s throughput per GiB of storage

$0.30/GiB-month*

* US N. Virginia

Provisioned Throughput mode pricing

Two pricing dimensions

Storage price

Pay only for the amount of storage you use per month

Includes 50 KiB/s throughput per GiB of storage

$0.30/GiB-month*

* US N. Virginia

Provisioned Throughput mode pricing

Two pricing dimensions

Storage price

Pay only for the amount of storage you use per month

Includes 50 KiB/s throughput per GiB of storage

$0.30/GiB-month*

Throughput price

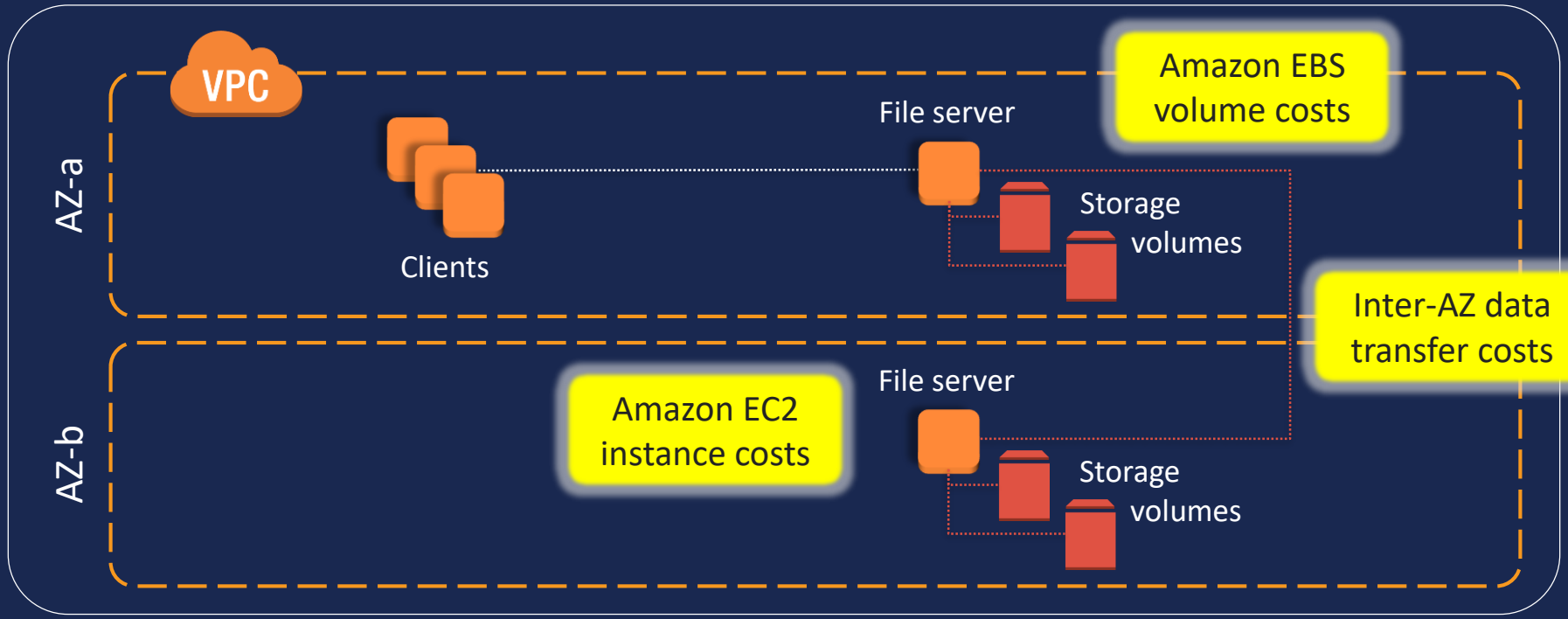
Pay only for the amount of throughput you provision per month

Above 50 KiB/s throughput per GiB of storage

$6.00/MiBps-month*

* US N. Virginia

Before Amazon EFS… DIY file storage costs



TCO example

For storing 500 GB, Amazon EFS is 70% less than DIY

Amazon EFS cost: (500 GB * $0.30/GB-month*) = $150 per month

For DIY, you might provision 600 GB of Amazon EBS (i.e., ~85% utilization):

Storage (2x 600 GB EBS gp2 volumes):	$120 per month
Compute (2x m4.xlarge instances):	$290 per month
Inter-AZ data transfer costs (est.):	$130 per month
Total	$540 per month

* US N. Virginia pricing

Designed for a wide spectrum of needs

Analytics

Media workflows

Enterprise apps and messaging

Web serving

Content management

Database backups

Container storage

Dev tooling

Home directories

Scale-out jobs

Metadata-intensive jobs

High throughput and parallel I/O

Low latency and serial I/O

Amazon EFS customers and partners

J.D. POWER
DigitalReLab
bambuser
zynga
ATLASSIAN
SPOKEO
Sysco
New York University
ASI DATA SCIENCE
webdam
Pinterest
infor
Cornell University
Seeking Alpha
here
BeSoftware
NETFLIX
Helix
zend
Disney
coursera
Celgene
custora
kissmetrics
ZS
THOMSON REUTERS
University of Pennsylvania
toast
Expedia
MONSANTO
FINRA
GE
TIBCO
workfront
PlanGrid
IBM
DELL Boomi

Phase 2:

Test and optimize

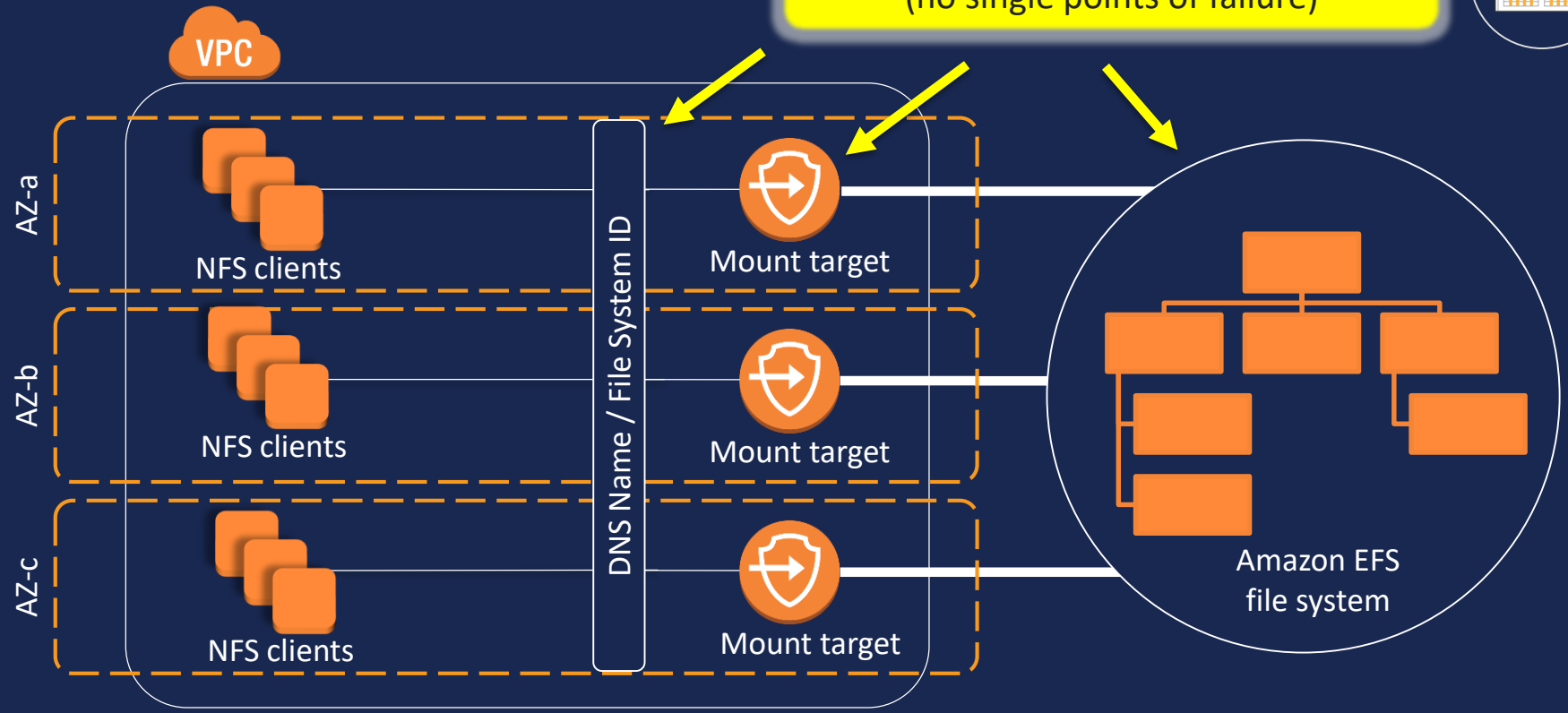
What do you think about?

Solution architecture

Functional testing

Performance testing and optimization

Amazon EFS architecture



How do I test Amazon EFS?

1. Create file system
2. Create mount targets
3. Mount file system

How to manage file systems

Create a file system

Create and manage mount targets

Tag a file system

Manage a file system

View details on a file system

Delete a file system

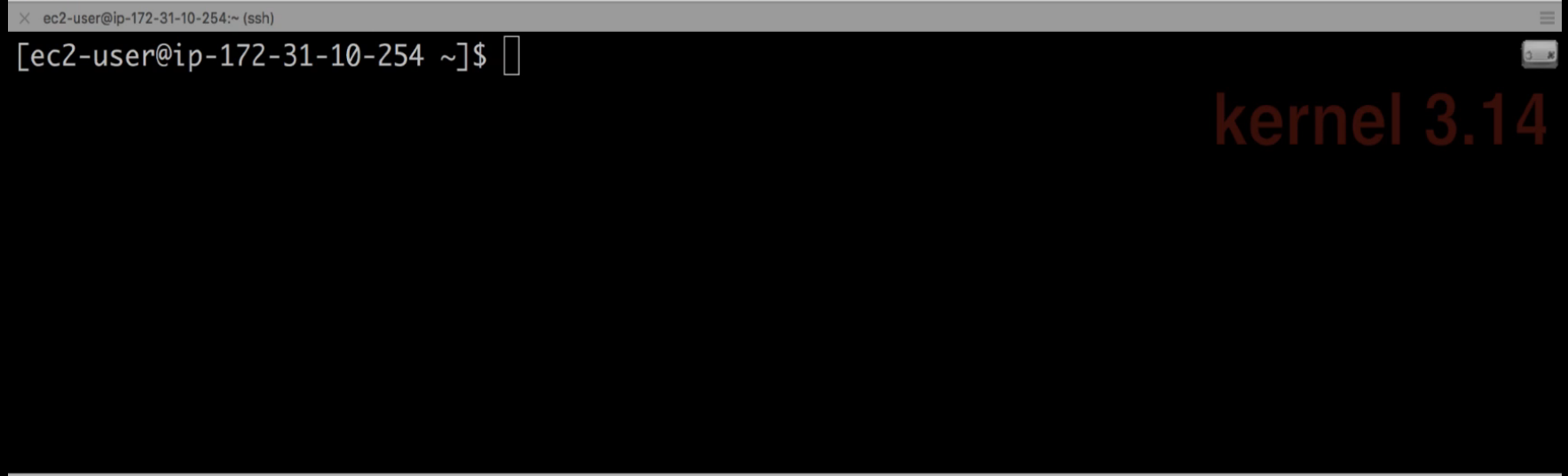
Kernel versions

Linux Kernel

use Linux kernel 4.0+*

e.g. Amazon Linux 17.09.1, Ubuntu16.04 or 17.10

* RHEL running Linux kernel 3.10+


ec2-user@ip-172-31-10-254:~ (ssh)
[ec2-user@ip-172-31-10-254 ~]$
kernel 3.14


ec2-user@ip-172-31-3-227:~
[ec2-user@ip-172-31-3-227 ~]$
kernel 4.14

Mount options

Recommended

NFS Mount Helper

Use for encrypted* or non-encrypted connections
* manual setup & configuration required

EFS Mount Helper

Use for encrypted or non-encrypted connections
Automatically uses recommended mount options

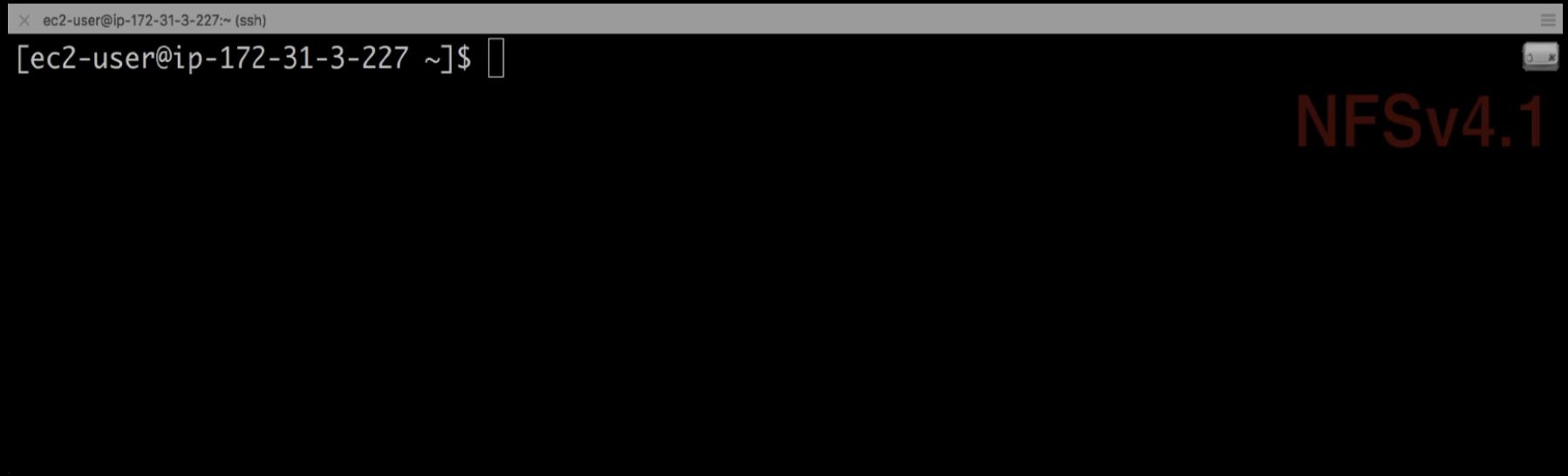
Standard NFS mount helper command:

```
mount -t nfs4 -o nfsvers=4.1,rsize=1048576,wsize=1048576,
hard,timeo=600,retrans=2 file-system-id.efs.region.amazonaws.com
efs-mount-point
```

Simple EFS mount helper command:

```
mount -t efs -o tls file-system-id efs-mount-point
```


ec2-user@ip-172-31-7-95:~
[ec2-user@ip-172-31-7-95 ~]$
NFSv4.0


ec2-user@ip-172-31-3-227:~ (ssh)
[ec2-user@ip-172-31-3-227 ~]$
NFSv4.1

Functional testing

POSIX compliant & compliant with NFS 4.0 & 4.1

But you always need to test your application

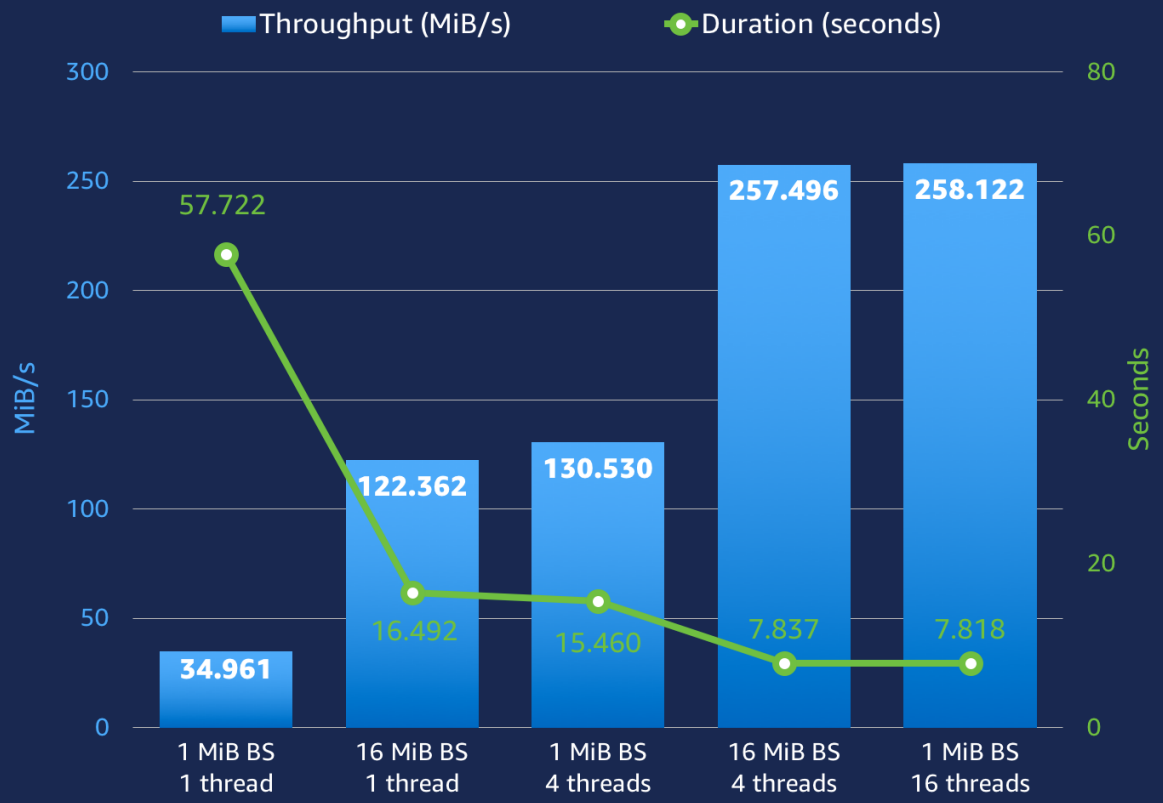
How do I test Amazon EFS performance?

Run a few performance tests against the file system

Hands-on Demo

aws Webinar

Throughput test results*

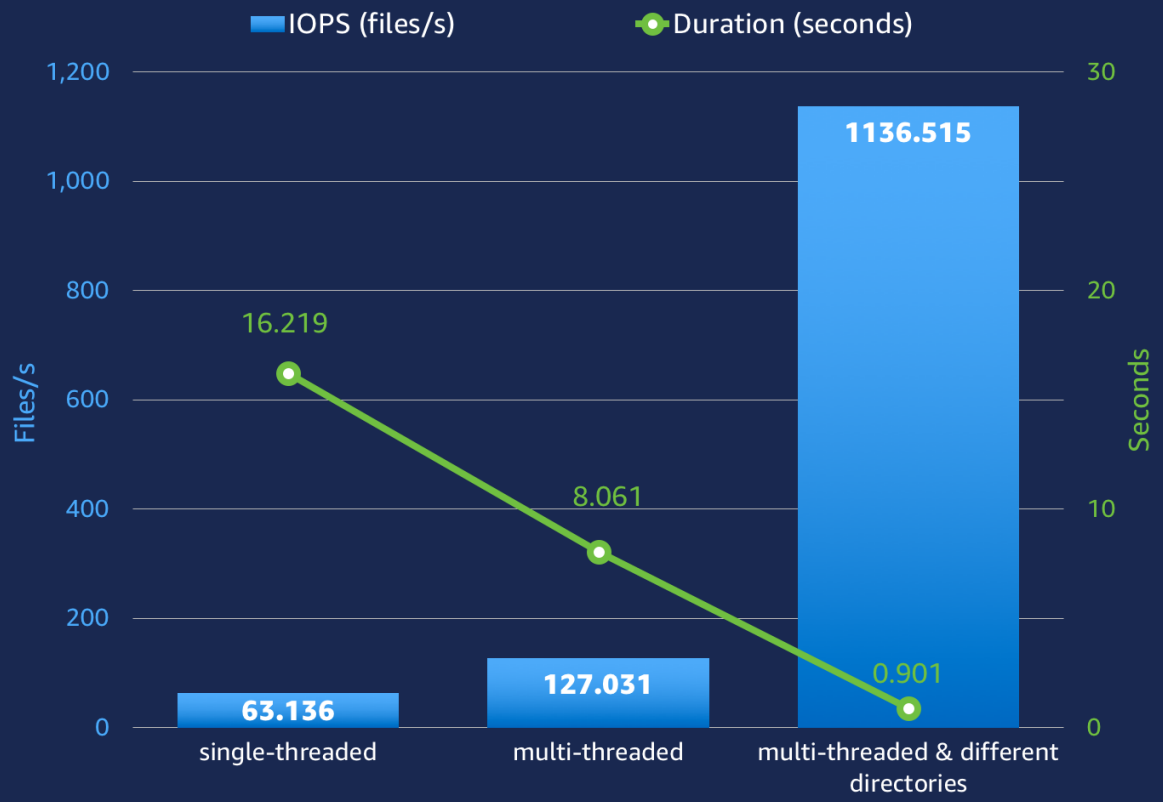


Best practices

- Use multiple threads
- Write to multiple directories in parallel
- Use larger IO size (aggregate IO)

*tested from a single c5.large instance

IOPS test results*



Best practices

- Use multiple threads
- Write to multiple directories in parallel

*tested from a single c5.large instance

Phase 3:
Ingest

What do you think about?

Where it’s coming from

How to move it as fast and easily as possible

Where is it coming from?

Corporate data center

Amazon Elastic Block Store

Other cloud file systems

Amazon Simple Storage Service

EFS File Sync

Sync data from existing file systems into Amazon EFS file systems

Simple
Set up and manage easily from the AWS Console

Fast
Up to 5x faster than standard Linux copy tools

Secure
Encrypted parallel data transfer to AWS

Use EFS File Sync to copy

- File systems from on-premises to EFS
- DIY in-cloud file systems to EFS
- EFS file systems between AWS Regions

How to leverage parallelism to copy data faster?

rsync	cp	fpsync	cp + GNU parallel	fpart + cpio + GNU parallel

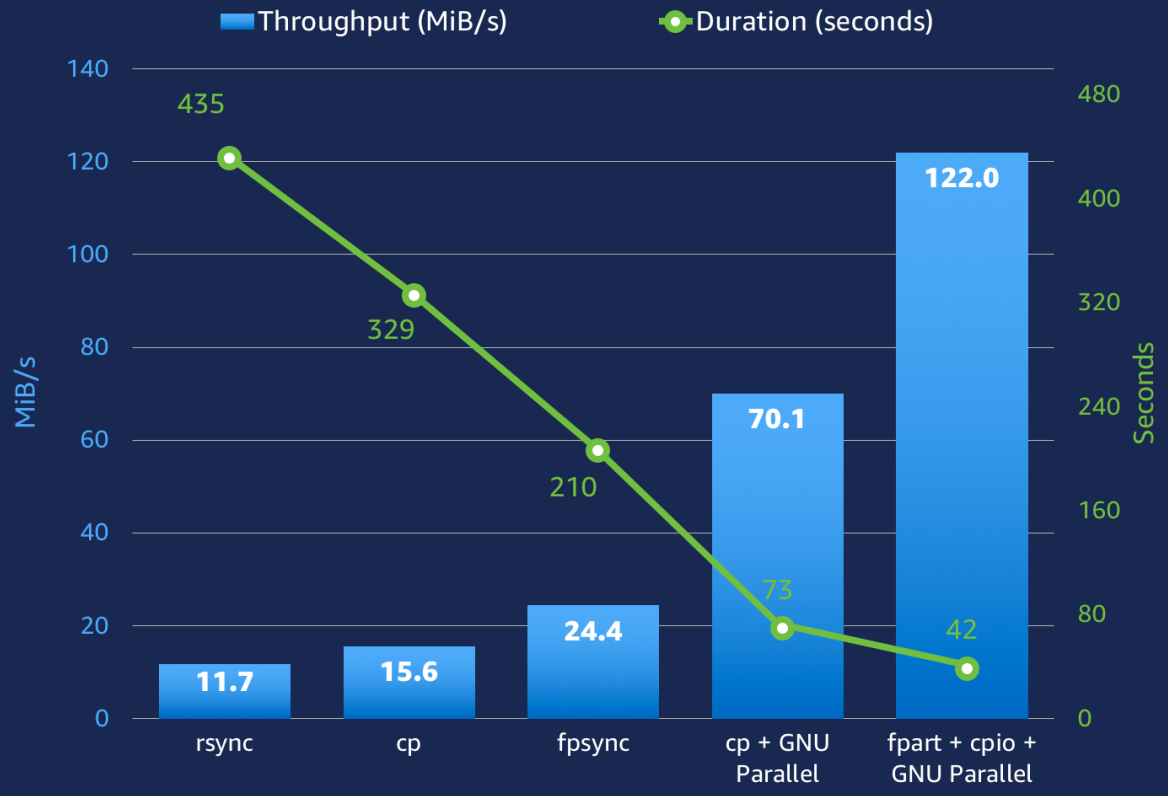
Hands-on demo

aws Webinar

How to leverage parallelism to copy data faster?

rsync	cp	fpsync	cp + GNU parallel	fpart + cpio + GNU parallel
single-threaded	single-threaded	multi-threaded	multi-threaded	multi-threaded
Poor (very chatty)	Fair	Good	Better	Best

File transfer test results*



Best Practices

- Use a multiple-threaded tool
- Use a less ”chatty” tool

*tested on a single c5.large instance

aws Webinar

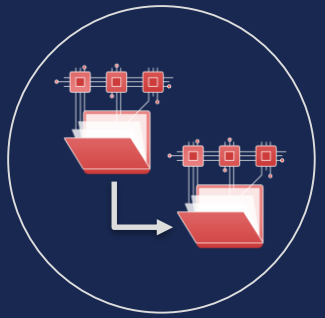
Phase 4:
Run

What do you think about? Not much.

View file system metrics

AWS CloudTrail API access logs

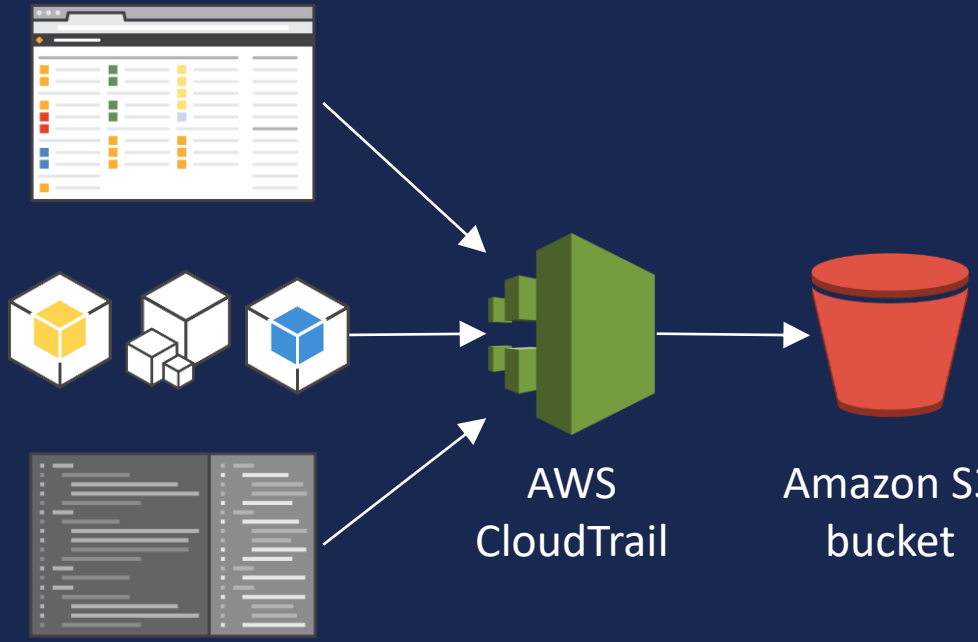
Amazon CloudWatch metrics



Perform backups

AWS CloudTrail API access logs

Logs EFS API calls from
Console, SDK, CLI



Amazon CloudWatch metrics

DataReadIOBytes

DataWriteIOBytes

MetaDataIOBytes

TotalIOBytes

BurstCreditBalance

PermittedThroughput

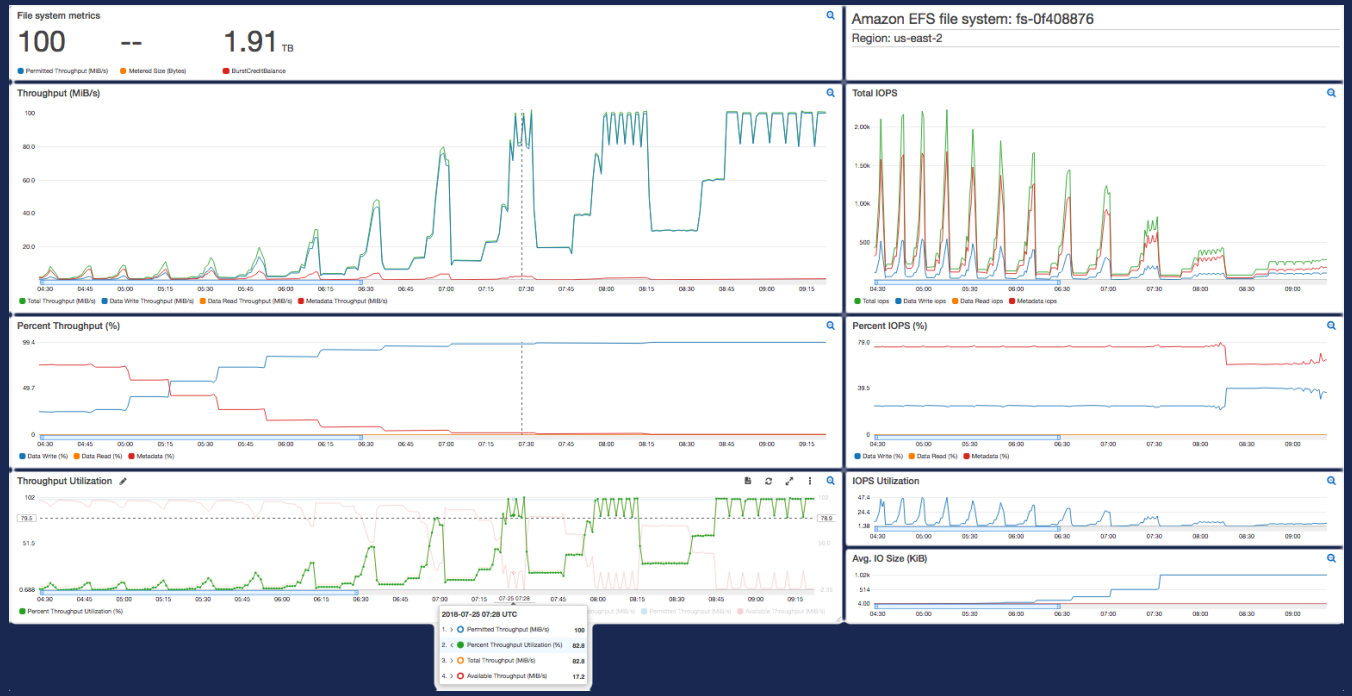
ClientConnections

PercentIOLimit*

Amazon CloudWatch

* General Purpose only

Amazon CloudWatch Dashboard



Hands-on demo

https://github.com/aws-samples/amazon-efs-tutorial/tree/master/monitoring

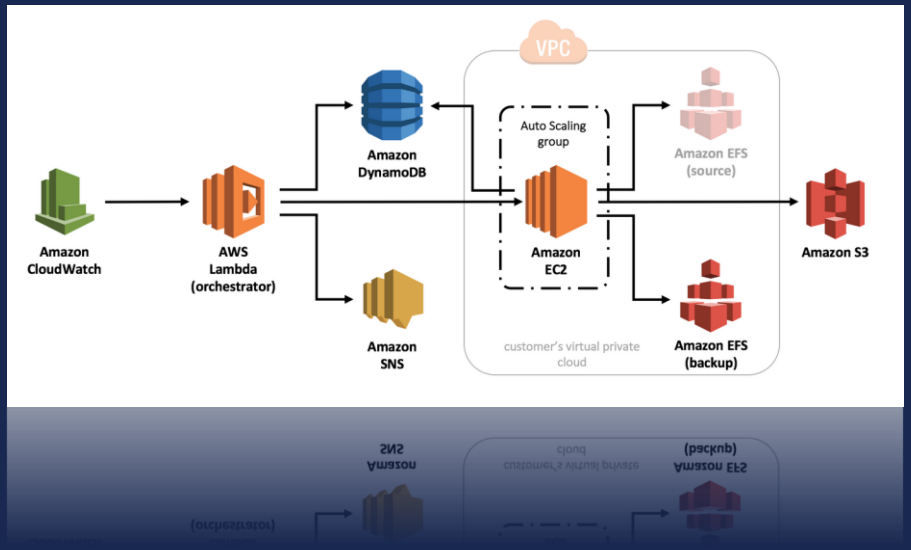
aws Webinar

Backup solution

EFS to EFS

Automatically backs up EFS

Easy to deploy



Wrapping up

Best practices

Test w/ General Purpose Performance Mode

Start w/ Bursting Throughput Mode

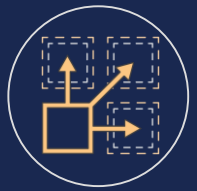
Linux kernel 4.0+

EFS mount helper (NFSv4.1)

Large IO size (aggregate IO)

Multiple threads

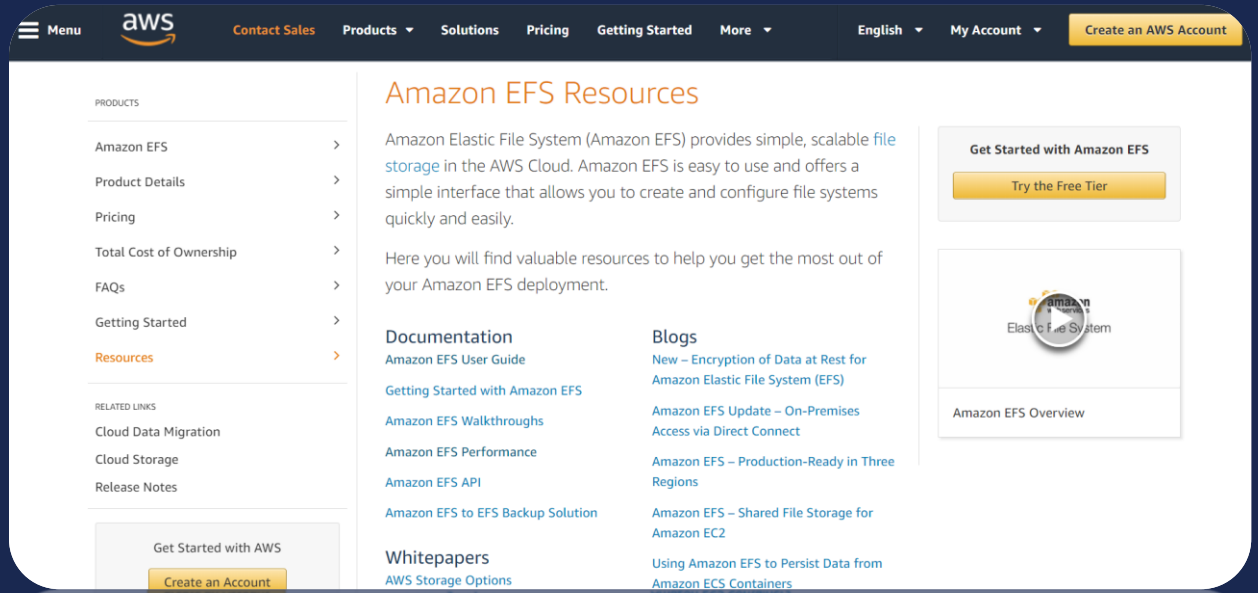
Multiple instances



Multiple directories

Monitor metrics

How to learn more: aws.amazon.com/efs



- Feature blogs
- Whitepapers
- Ref architectures
- TCO calculator
- 10-minute tutorials
- Documentation

Amazon EFS tutorials

https://github.com/aws-samples/amazon-efs-tutorial

README.md

Amazon Elastic File System (Amazon EFS)

Tutorials

Version 1.0.1

Thank you

Darryl S. Osborne
Solutions Architect – Amazon File Services
darrylo@amazon.com

Q & A